

Soft Decoding of a Qubit Readout Apparatus

B. D'Anjou

Department of Physics, McGill University, Montreal, Quebec, H3A 2T8, Canada

W.A. Coish

*Department of Physics, McGill University, Montreal, Quebec, H3A 2T8, Canada and
Canadian Institute for Advanced Research, Toronto, Ontario, M5G 1Z8, Canada*

(Dated: December 25, 2014)

Qubit readout is commonly performed by thresholding a collection of analog detector signals to obtain a sequence of single-shot bit values. The intrinsic irreversibility of the mapping from analog to digital signals discards soft information associated with an *a posteriori* confidence that can be assigned to each bit value when a detector is well characterized. Accounting for soft information, we show significant improvements in enhanced state detection with the quantum repetition code as well as quantum state or parameter estimation. These advantages persist in spite of non-Gaussian features of realistic readout models, experimentally relevant small numbers of qubits, and finite encoding errors. These results show useful and achievable advantages for a wide range of current experiments on quantum state tomography, parameter estimation, and qubit readout.

PACS numbers: 03.65.Ta, 03.67.Ac, 03.65.Wj

In most quantum-measurement tasks, the goal is to extract information encoded in a stream of quantum states. For a single-qubit expectation value, the stream is a collection of identically prepared single qubits. To perform a Bell-inequality measurement, the stream is a collection of entangled qubit pairs. For quantum error detection or correction, the stream consists of many qubits that make up the code, on which multiqubit syndrome measurements are performed. In practice, in all of these scenarios, information is commonly extracted by measuring individual qubits or joint observables in a single shot [1–8]. While this strategy can be optimal for extracting a single bit of information, e.g., the state $|\pm\rangle$ of a single qubit, it is generally suboptimal when considering streams of data. A single-shot qubit readout typically involves the irreversible conversion of an analog outcome \mathcal{O} from a readout apparatus (e.g., a current or voltage pulse, the quadrature of a microwave tone, etc.) into a binary outcome c_{\pm} via thresholding [9–17] (see Fig. 1). Thresholding erases information about the posterior probability $P(\pm|\mathcal{O})$ that can be ascribed to each bit value given \mathcal{O} . In contrast, when a sequence of analog readout outcomes \mathcal{O} is fed into a decoder that accepts analog values as input, the frequency of decoding errors can be significantly reduced [18]. Such *soft-decoding* techniques have been central to the development of capacity-achieving classical codes [19] now used in deep-space communications and high-bandwidth 3G/4G cellular networks. Soft-decision decoding has been applied to quantum codes [20, 21] and to schemes for fault-tolerant quantum computing [22], in which the soft decision is made by correlating multiple single-shot qubit readout outcomes. Soft decoding has also been identified as an important tool for continuous-variable quantum key distribution [23].

While soft-decoding methods are routinely applied to

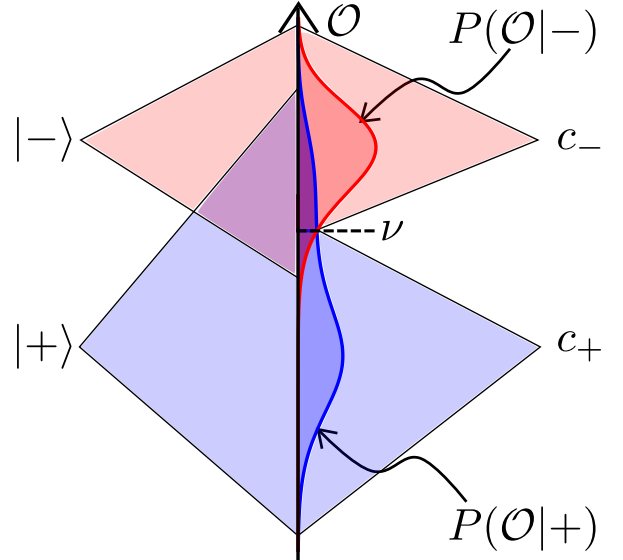


FIG. 1. (Color online) The single-shot readout thresholding procedure (threshold ν) erases information by irreversibly converting a continuum of analog readout outcomes \mathcal{O} into a single binary value c_{\pm} . The conditional probability distributions $P(\mathcal{O}|\pm)$ are typical of the readout performed in Refs. [11, 24] and described in Ref. [25].

classical noisy signals for communication applications, they have not seen widespread application in qubit readout methods. Here, we exploit the fact that the qubit readout itself can be treated as a communication channel characterized by a pair of conditional probability distributions $P(\mathcal{O}|\pm)$ for the analog signal \mathcal{O} (see Fig. 1), *even if* the readout apparatus is designed to perform a binary measurement. Soft decoding of a readout appa-

ratus can lead to significant improvements in a number of quantum-information tasks. To achieve these improvements, the physical characteristics of the readout dynamics and the noise must be well understood to determine the distributions $P(\mathcal{O}|\pm)$. Readout errors are especially sensitive to the tails of these distributions, so it is important to understand non-Gaussian features (fat tails or bimodality) to reap the benefits of soft decoding. Crucially, the distributions $P(\mathcal{O}|\pm)$, which must already be known to characterize the single-shot readout fidelity, can be measured or modeled accurately for several state-of-the-art qubit implementations [9–17, 25–28].

In this Letter, we explicitly demonstrate the advantages of soft decoding through two experimentally significant examples: enhanced state detection via the quantum repetition code, and state or parameter estimation. In particular, the number of qubits required for efficient enhanced state detection can be reduced, through data processing alone, by up to a factor of 2, an advantage which persists for a small number of qubits and finite encoding errors. Because additional qubits are an expensive resource, this result is of immediate practical importance. We also extend a result of Ref. [29] by making use of soft information to further improve the precision of measurements of Pauli operators required for state tomography. In both cases, we benchmark the improvement by comparing the performance of the widely used and efficient maximum-likelihood estimation [30] when applied to analog instead of thresholded qubit readout outcomes. Crucially, we find and characterize significant improvements not only for the idealized Gaussian readout, but also for the realistic non-Gaussian readout investigated in Ref. [25] and relevant to many experiments [11, 14, 24, 26].

Enhanced state detection.— In the quantum repetition code, a logical qubit (with basis $\{|0\rangle, |1\rangle\}$) in the state $|\psi\rangle = \alpha_0|0\rangle + \alpha_1|1\rangle$ is encoded into N physical qubits (with basis $\{|+\rangle, |-\rangle\}$): $|\psi_N\rangle = \alpha_0|-\rangle^{\otimes N} + \alpha_1|+\rangle^{\otimes N}$ [31–33]. The redundant outcomes of independent measurements of the N physical qubits are then correlated to reduce measurement errors. The simplest approach is to measure each qubit in a single shot and assign the binary value $c_i = c_\pm$ to the i th qubit. When the encoded states $|0\rangle$ and $|1\rangle$ are equally likely *a priori*, the optimal approach is to calculate the likelihood ratio [34] of the data $\{c_i\}$:

$$\Lambda_c \equiv \frac{P(\{c_i\}|1)}{P(\{c_i\}|0)} = \prod_{i=1}^N \frac{P(c_i|+)}{P(c_i|-)}. \quad (1)$$

In Eq. (1), $P(c_i|\pm)$ is the probability to obtain the value c_i given that the i th qubit is in the state $|\pm\rangle$. The projected state of the logical qubit is most likely $|1\rangle$ ($|0\rangle$) if $\Lambda_c > 1$ ($\Lambda_c < 1$). We may rewrite the likelihood ratio, Eq. (1), in terms of the conditional single-shot error rates

$$\epsilon_\pm \equiv P(c_\mp|\pm) < 1/2:$$

$$\Lambda_c = \left(\frac{1 - \epsilon_+}{\epsilon_-}\right)^{n_+} \cdot \left(\frac{\epsilon_+}{1 - \epsilon_-}\right)^{N - n_+}, \quad (2)$$

where n_+ is the number of measurements for which the outcome c_+ occurred. For a binary symmetric readout, $\epsilon = \epsilon_\pm$, Eq. (2) results in a simple majority vote since $n_+ > N/2$ ($n_+ < N/2$) implies $\Lambda_c > 1$ ($\Lambda_c < 1$).

Equation (2) is a maximum-likelihood estimator applied to single-shot readout outcomes. However, a physical readout apparatus typically yields an analog readout outcome \mathcal{O}_i that need not be thresholded to a binary value c_i . The observable \mathcal{O}_i could be, for example, the time average of a fluorescence signal [9, 13, 33], the peak of a current pulse through a single-electron transistor or quantum point contact [11, 24, 26], the quadrature of a microwave tone [15, 16, 27], or even the likelihood ratio of a single-shot readout [9, 17, 25, 28, 35]. Thresholding leads to an irreversible loss of information about the confidence $P(\pm|\mathcal{O}_i)$ in each bit value. Soft decoding, which makes full use of that information, is achieved by instead applying the maximum-likelihood estimator to the analog readout outcomes:

$$\Lambda_{\mathcal{O}} \equiv \frac{P(\{\mathcal{O}_i\}|1)}{P(\{\mathcal{O}_i\}|0)} = \prod_{i=1}^N \frac{P(\mathcal{O}_i|+)}{P(\mathcal{O}_i|-)}, \quad (3)$$

where $P(\mathcal{O}_i|\pm)$ is the probability density for outcome \mathcal{O}_i given that the i th qubit is in the state $|\pm\rangle$.

To take full advantage of soft decoding, it is necessary to have an accurate representation of the conditional probability distributions $P(\mathcal{O}|\pm)$ for the analog qubit readout outcomes. A common idealization for a readout is the Gaussian readout, $P(\mathcal{O}|\pm) = \sqrt{r/2\pi} \exp[-(\mathcal{O} \mp 1)^2 r/2]$, where r is the power signal-to-noise ratio. Soft decoding of the Gaussian readout with maximum-likelihood estimators such as Eqs. (1) and (3) has been extensively studied in the context of classical communication theory [18, 34]. Since a projective measurement collapses $|\psi_N\rangle$ to either $|+\rangle^{\otimes N}$ or $|-\rangle^{\otimes N}$, the advantage obtained by soft decoding of the readout apparatus translates directly to the quantum case. More precisely, for $r \gg 1$, the number of qubits N_c and $N_{\mathcal{O}}$ required to achieve a target error rate ε using Λ_c and $\Lambda_{\mathcal{O}}$, respectively, are related (for N_c odd) by [36]:

$$N_{\mathcal{O}} = \frac{N_c + 1}{2} + \frac{N_c - 1}{2} \frac{\ln r}{r} + O\left(\frac{N_c}{r}\right). \quad (4)$$

Thus, soft decoding can reduce the number of required physical qubits by up to a factor of 2 compared to the majority vote (asymptotically, $N_c \sim 2N_{\mathcal{O}}$, or alternatively, for fixed N , $\varepsilon_{\mathcal{O}} \sim \varepsilon_c^2$ up to a logarithmic prefactor in ε_c). Intuitively, this advantage arises since the majority vote ignores all information contained in strings for which more than half of the bits are corrupted, while

soft decoding associates every string with some confidence. A similar asymptotic advantage exists for arbitrary block codes transmitted through a Gaussian communication channel [18]. Importantly, Eq. (4) is valid for the regime of reasonably small N relevant to recent experiments [2, 5–8, 37, 38]. Moreover, the form of the subleading corrections in Eq. (4) suggests that they can be small for realistic experimental values of r . We have indeed verified, using the exact analytical expressions for the error rates [36], that an advantage persists for low signal-to-noise ratio and relatively small N . For example, we require $N_{\mathcal{O}} = 6$ instead of $N_c = 9$ to reach an error rate $\varepsilon < 3 \times 10^{-4}$ for $r = 2$.

Realistic qubit readouts are typically not well represented by Gaussian probability distributions [9, 11, 13, 25, 28]. To verify that soft decoding of the readout apparatus still provides an advantage in experimentally relevant cases, we apply the estimators in Eqs. (1) and (3) to the realistic non-Gaussian “peak-signal” readout implemented in Refs. [11, 24] and for which the distributions $P(\mathcal{O}|\pm)$ were analyzed in Ref. [25]. In this measurement, the analog outcome \mathcal{O} is the peak value of a finite-duration current pulse signalling the excited state $|+\rangle$ and subject to Gaussian white noise (see Ref. [36] for a summary). A typical pair of distributions for this readout is illustrated in Fig. 1. Even though the distribution $P(\mathcal{O}|+)$ has strong non-Gaussian features, soft decoding still gives an appreciable advantage. For example, Monte Carlo simulations with 10^6 random records show that, similar to the Gaussian readout, we require $N_{\mathcal{O}} = 6$ instead of $N_c = 9$ to reach an error rate $\varepsilon < 0.05$ for a signal-to-noise ratio $r = 2$ [36].

To account for errors during encoding, we allow for uncorrelated bit flips with probability η for both states $|\pm\rangle$. The probability distributions for the analog readout outcomes then become $P(\mathcal{O}_i|1/0) = (1 - \eta)P(\mathcal{O}_i|+/-) + \eta P(\mathcal{O}_i|-/+)$, giving a modified version of the likelihood ratio, Eq. (3). We find that when encoding errors η are sufficiently large, the soft-decoding procedure reduces to a simple thresholding procedure [36]. However, for the Gaussian readout, soft decoding can still give an advantage over thresholding if

$$\eta \lesssim e^{-2r} \quad (5)$$

when $r \gg 1$. Thus, the encoding bit-flip rate must merely be smaller than some power of the single-shot readout error rate ϵ ($\epsilon \sim e^{-\frac{r}{2}}$ up to logarithmic corrections). To verify this, we have performed a Monte Carlo simulation of the error rate for the Gaussian readout by generating 10^7 random measurement records taking into account the bit-flip rate η [36]. For example, we find that for $r = 2$ and $\eta = 1\%$, we require $N_{\mathcal{O}} = 6$ instead of $N_c = 9$ qubits to achieve $\varepsilon < 8 \times 10^{-4}$. Similarly, for the peak-signal readout described in Ref. [25], we find from a simulation of 10^6 random measurement records that for a signal-to-noise ratio of $r = 2$ and $\eta = 5\%$, we require $N_{\mathcal{O}} = 6$

instead of $N_c = 9$ qubits to achieve $\varepsilon < 0.08$ [36].

State and parameter estimation.— Many quantum information processing applications, such as state and process tomography [1, 4, 29, 39, 40] and parameter estimation [41], benefit from accurate and precise estimation of qubit observables (e.g., the Pauli operators). Analog data processing has been used extensively, e.g., for parameter [42] and state [43] estimation in quantum optical systems, where it is often natural to process quasicontinuous field quadratures or photon counts. For many qubit systems, the common approach is instead to threshold the data. Thresholding the data is generally suboptimal, as we now illustrate.

For definiteness, we consider estimating the quantum expectation value $s_0 = \langle \sigma_z \rangle$ of the single-qubit Pauli operator σ_z (in the basis $|\pm\rangle$) from the independent readout of N identically prepared copies of a qubit. As in the case of the repetition code, we compare the standard maximum-likelihood estimator (MLE) [30] applied to the analog data set $\{\mathcal{O}_i\}$ instead of the thresholded data set $\{c_i\}$ in order to benchmark the improvement. In both cases, the MLE is the value s that maximizes the likelihood function $\mathcal{L}(s) = \prod_{i=1}^N P(\mathcal{O}_i/c_i|s)$ under the constraint $-1 \leq s \leq +1$. In practice, the MLE is obtained by maximizing the equivalent log-likelihood function $\ell(s) = N^{-1} \ln \mathcal{L}(s)$. The MLE is asymptotically unbiased, normally distributed, and minimizes the variance [i.e. saturates the Cramér-Rao bound, see Eq. (7), below] for large N [30].

When the analog data are thresholded, the MLE is the (bias-corrected) thresholded average $s_{\text{TA}} = N^{-1} \sum_{i=1}^N c_i$ considered, e.g., in Ref. [29]. This estimate does not make use of the soft information contained in the distributions $P(\mathcal{O}|\pm)$ for reconstruction of s_0 . In contrast, the soft-decoded estimate s_{SD} obtained by applying the MLE to the analog data set makes full use of the distributions $P(\mathcal{O}|\pm)$. In Ref. [29], the alternative *soft average* $s_{\text{SA}} = N^{-1} \sum_{i=1}^N \mathcal{O}_i$ was also employed as an estimator for s_0 , but this approach is also suboptimal [36].

We will measure the deviation of an estimate s from the true value s_0 with the mean squared error (MSE) ζ , given by the sum of the variance and of the squared bias of the estimator, $\zeta \equiv \langle (s - \langle s \rangle)^2 \rangle + (\langle s \rangle - s_0)^2$. Here, the statistical average $\langle \rangle$ is taken with respect to the distribution of outcomes:

$$P(\mathcal{O}/c|s_0) = \frac{1+s_0}{2} P(\mathcal{O}/c|+) + \frac{1-s_0}{2} P(\mathcal{O}/c|-). \quad (6)$$

For this distribution, $\ell(s)$ is a concave function with a unique maximum. A general expression for the asymptotic MSE of the thresholded average s_{TA} can be derived [36]. For the Gaussian readout, it takes the simple form reported in Ref. [29], $\zeta_{\text{TA}} \sim [(1-2\epsilon)^{-2} - s_0^2]/N$. For the soft-decoded MLE estimate, the asymptotic MSE can be computed directly from the Fisher information of

$P(\mathcal{O}|s_0)$:

$$\zeta_{\text{SD}} \sim -\frac{1}{N} \left\langle \left\langle \frac{\partial^2 \ln P(\mathcal{O}|s_0)}{\partial s_0^2} \right\rangle \right\rangle^{-1}. \quad (7)$$

Here, we use the symbol “ \sim ” to indicate a strict asymptotic equality. From Eqs. (6) and (7), an explicit asymptotic form for ζ_{SD} can be found in terms of the distributions $P(\mathcal{O}|\pm)$ [36]:

$$\zeta_{\text{SD}} \sim \frac{1}{N} \cdot \frac{1 - s_0^2}{1 - I}, \quad I = \int d\mathcal{O} \frac{P(\mathcal{O}|+)P(\mathcal{O}|-)}{P(\mathcal{O}|s_0)}. \quad (8)$$

In Eq. (8), the integral I contains all information about the noise introduced by the readout apparatus. The remaining contribution when $I = 0$ is the quantum shot noise (projection noise), which reflects the choice of a particular measurement basis. Since I has the form of an overlap integral, it is especially important to understand the tails of the (generally non-Gaussian) readout distributions $P(\mathcal{O}|\pm)$.

To quantitatively verify that soft decoding can improve state estimation, we set $s_0 = 0$ and evaluate Eq. (8) numerically for both the Gaussian readout and the peak-signal readout of Ref. [25]. We plot the asymptotic MSE as a function of the signal-to-noise ratio r in Fig. 2. We also plot the asymptotic MSE of the bias-corrected thresholded average, ζ_{TA} . Figure 2 confirms that soft decoding always outperforms thresholding (i.e., $\zeta_{\text{SD}} < \zeta_{\text{TA}}$). As r increases, ζ_{TA} and ζ_{SD} exhibit an approximate power-law approach to the projection-noise limit for the (non-Gaussian) peak-signal readout of Ref. [25], whereas they decrease exponentially for the Gaussian readout. In the intermediate regime for r , there is a clear advantage in soft decoding with the MLE for both readouts, demonstrating substantial benefits in the experimentally relevant regime of signal-to-noise ratios, $r \sim 1$ for the Gaussian readout [29], and $r \sim 10$ for the peak-signal readout.

To show that the asymptotic advantage persists when N is finite, we calculate ζ from a Monte Carlo simulation with $N = 100$. We first randomly generate 5×10^4 measurement records $\{\mathcal{O}_i\}$ from the distribution $\prod_{i=1}^N P(\mathcal{O}_i|s_0)$ for both the Gaussian readout and non-Gaussian peak-signal readout of Ref. [25]. For each measurement record, we calculate s_{TA} and optimize the log-likelihood function $\ell(s)$ to obtain s_{SD} . We then directly obtain ζ from the variance and bias of the simulated estimates. The results are displayed (open symbols) in Fig. 2 for $s_0 = 0$. The simulated data points coincide with the asymptotic predictions.

In conclusion, we have shown that making use of the soft information contained in the analog outcomes of a qubit readout apparatus, as opposed to irreversibly thresholding each qubit to a binary value, can significantly improve the performance of quantum information processing tasks involving the measurement of many

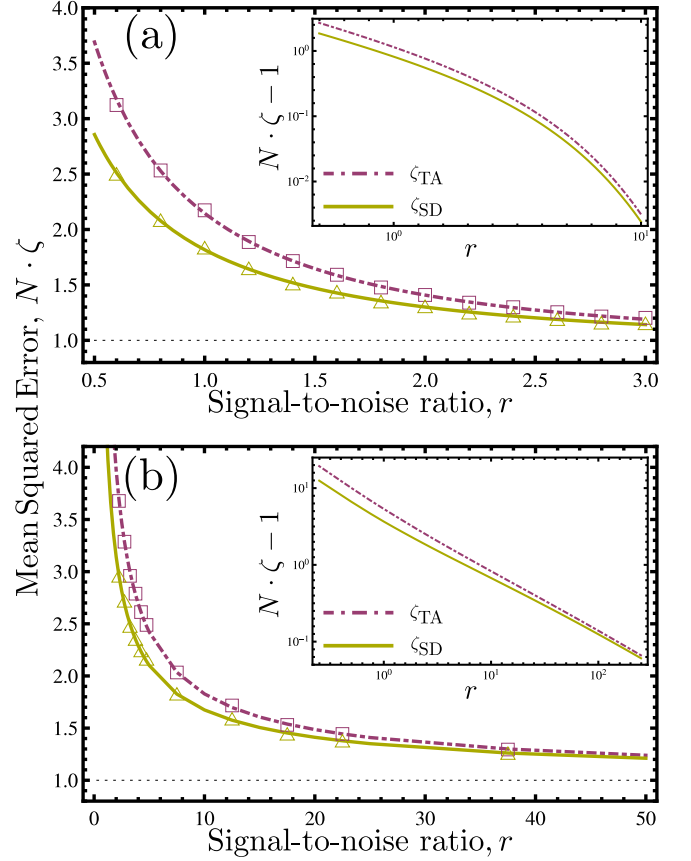


FIG. 2. (Color online) Asymptotic normalized MSEs $N \cdot \zeta$ of the soft-decoded estimate (solid gold), Eq. (8), and of the thresholded average (dot-dashed magenta line), given in Ref. [36], as a function of the signal-to-noise ratio r assuming $s_0 = \langle \sigma_z \rangle = 0$ for (a) the Gaussian readout and (b) the “peak-signal” readout of Ref. [25]. The finite- N MSEs of the thresholded average (magenta squares) and soft-decoded estimate (gold triangles) are obtained from 5×10^4 randomly generated measurement records of $N = 100$ qubits. Insets: Asymptotic MSEs on a logarithmic scale.

qubits. We have focused on two examples of practical importance. In the case of enhanced state detection with the quantum repetition code, the number of qubits required to achieve a given error rate can be reduced by up to a factor of 2 through improved data processing alone. Importantly, we have shown that an advantage persists for small numbers of qubits and finite encoding errors. In addition, we have shown that optimal processing of analog qubit readout outcomes can appreciably increase the precision on the measurement of qubit observables (e.g., the Pauli operators). Crucially, in both cases we have demonstrated a significant improvement for an experimentally relevant, non-Gaussian qubit readout model [11, 14, 24–26].

Our results offer encouraging prospects for direct improvements of both small and large scale quantum information processing applications through soft decoding of

the qubit readout. For example, readout error models for decoders of topological codes [21, 44] could be modified using the ideas presented here to accept analog readout outcomes with realistic statistics at the single-physical-qubit level, improving error detection rates. While there are many possible extensions of this work, the direct improvements we have shown to enhanced state detection and quantum state or parameter estimation are both practical and immediately realizable in a wide array of current experiments.

We thank L. Childress, A. Fowler, and D. Poulin for useful discussions. We acknowledge financial support from the National Sciences and Engineering Research Council of Canada (NSERC), the Canadian Institute for Advanced Research (CIFAR), the Fonds de la recherche en nature et technologies (FRQNT), and the Institut Transdisciplinaire d'Information Quantique (INTRIQ).

-
- [1] H. Häffner, W. Hänsel, C. Roos, J. Benhelm, M. Chwalla, T. Körber, U. Rapol, M. Riebe, P. Schmidt, C. Becher, *et al.*, *Nature* **438**, 643 (2005).
 - [2] P. Schindler, J. T. Barreiro, T. Monz, V. Nebendahl, D. Nigg, M. Chwalla, M. Hennrich, and R. Blatt, *Science* **332**, 1059 (2011).
 - [3] H. Bernien, B. Hensen, W. Pfaff, G. Koolstra, M. Blok, L. Robledo, T. Taminiau, M. Markham, D. Twitchen, L. Childress, *et al.*, *Nature* **497**, 86 (2013).
 - [4] J. M. Chow, A. D. Córcoles, J. M. Gambetta, C. Rigetti, B. R. Johnson, J. A. Smolin, J. R. Rozen, G. A. Keefe, M. B. Rothwell, M. B. Ketchen, and M. Steffen, *Phys. Rev. Lett.* **107**, 080502 (2011).
 - [5] M. Reed, L. DiCarlo, S. Nigg, L. Sun, L. Frunzio, S. Girvin, and R. Schoelkopf, *Nature* **482**, 382 (2012).
 - [6] J. M. Chow, J. M. Gambetta, E. Magesan, S. J. Srinivasan, A. W. Cross, D. W. Abraham, N. A. Masluk, B. Johnson, C. A. Ryan, and M. Steffen, *arXiv:1311.6330* (2013).
 - [7] O.-P. Saira, J. P. Groen, J. Cramer, M. Meretska, G. de Lange, and L. DiCarlo, *Phys. Rev. Lett.* **112**, 070502 (2014).
 - [8] R. Barends, J. Kelly, A. Megrant, A. Veitia, D. Sank, E. Jeffrey, T. White, J. Mutus, A. Fowler, B. Campbell, *et al.*, *Nature* **508**, 500 (2014).
 - [9] A. H. Myerson, D. J. Szwer, S. C. Webster, D. T. C. Allcock, M. J. Curtis, G. Imreh, J. A. Sherman, D. N. Stacey, A. M. Steane, and D. M. Lucas, *Phys. Rev. Lett.* **100**, 200502 (2008).
 - [10] C. Barthel, D. J. Reilly, C. M. Marcus, M. P. Hanson, and A. C. Gossard, *Phys. Rev. Lett.* **103**, 160503 (2009).
 - [11] A. Morello, J. J. Pla, F. A. Zwanenbourg, K. W. Chan, K. Y. Tan, H. Huebl, M. Möttönen, C. D. Nugroho, C. Yang, J. A. van Donkelaar, *et al.*, *Nature* **467**, 687 (2010).
 - [12] P. Neumann, J. Beck, M. Steiner, F. Rempp, H. Fedder, P. R. Hemmer, J. Wrachtrup, and F. Jelezko, *Science* **329**, 542 (2010).
 - [13] L. Robledo, L. Childress, H. Bernien, B. Hensen, P. F. Alkemade, and R. Hanson, *Nature* **477**, 574 (2011).
 - [14] J. J. Pla, K. Y. Tan, J. P. Dehollain, W. H. Lim, J. J. Morton, F. A. Zwanenbourg, D. N. Jamieson, A. S. Dzurak, and A. Morello, *Nature* **496**, 334 (2013).
 - [15] Z. R. Lin, K. Inomata, W. D. Oliver, K. Koshino, Y. Nakamura, J. S. Tsai, and T. Yamamoto, *Appl. Phys. Lett.* **103**, 132602 (2013).
 - [16] Y. Liu, S. Srinivasan, D. Hover, S. Zhu, R. McDermott, and A. Houck, *arXiv:1401.5184* (2014).
 - [17] T. Harty, D. Allcock, C. Ballance, L. Guidoni, H. Janacek, N. Linke, D. Stacey, and D. Lucas, *arXiv:1403.1524* (2014).
 - [18] D. Chase, *Information Theory, IEEE Transactions on* **18**, 170 (1972).
 - [19] E. Guizzo, “Closing in on the perfect code,” <http://spectrum.ieee.org/computing/software/closing-in-on-the-perfect-code> (2004).
 - [20] D. Poulin, *Phys. Rev. A* **74**, 052333 (2006).
 - [21] G. Duclos-Cianci and D. Poulin, in *Information Theory Workshop (ITW), 2010 IEEE* (IEEE, 2010) pp. 1–5.
 - [22] H. Goto and H. Uchikawa, *Scientific reports* **3** (2013).
 - [23] M. Mondin, F. Daneshgaran, M. Delgado, and F. Mesiti, in *Personal Satellite Services* (Springer, 2010) pp. 305–316.
 - [24] J. M. Elzerman, R. Hanson, L. H. W. Van Beveren, B. Witkamp, L. M. K. Vandersypen, and L. P. Kouwenhoven, *Nature* **430**, 431 (2004).
 - [25] B. D’Anjou and W. A. Coish, *Phys. Rev. A* **89**, 012313 (2014).
 - [26] M. Veldhorst, J. C. C. Hwang, C. H. Yang, A. W. Leenstra, B. de Ronde, J. P. Dehollain, J. T. Muhonen, F. E. Hudson, K. M. Itoh, A. Morello, and A. S. Dzurak, *arXiv preprint arXiv:1407.1950* (2014).
 - [27] E. Jeffrey, D. Sank, J. Y. Mutus, T. C. White, J. Kelly, R. Barends, Y. Chen, Z. Chen, B. Chiaro, A. Dunsworth, A. Megrant, P. J. J. O’Malley, C. Neill, P. Roushan, A. Vainsencher, J. Wenner, A. N. Cleland, and J. M. Martinis, *Phys. Rev. Lett.* **112**, 190504 (2014).
 - [28] J. Gambetta, W. A. Braff, A. Wallraff, S. M. Girvin, and R. J. Schoelkopf, *Phys. Rev. A* **76**, 012325 (2007).
 - [29] C. A. Ryan, B. R. Johnson, J. M. Gambetta, J. M. Chow, M. P. da Silva, O. E. Dial, and T. A. Ohki, *arXiv:1310.6448* (2013).
 - [30] H. Cramér, *Mathematical methods of statistics* (Princeton University Press, Princeton, NJ, 1946) Chap. 32–33.
 - [31] P. Deuar and W. J. Munro, *Phys. Rev. A* **61**, 010306 (1999).
 - [32] D. P. DiVincenzo, in *Scalable Quantum Computers: Paving the Way to Realization*, edited by S. L. Braunstein, H.-K. Lo, and P. Kok (Wiley-VCH, Berlin, Germany, 2001) Chap. 1, pp. 1–13.
 - [33] T. Schaetz, M. D. Barrett, D. Leibfried, J. Britton, J. Chiaverini, W. M. Itano, J. D. Jost, E. Knill, C. Langer, and D. J. Wineland, *Phys. Rev. Lett.* **94**, 010501 (2005).
 - [34] J. M. Wozencraft and I. M. Jacobs, *Principles of communication engineering* (John Wiley & Sons, New York, U.S.A., 1965) Chap. 4.
 - [35] D. B. Hume, T. Rosenband, and D. J. Wineland, *Phys. Rev. Lett.* **99**, 120502 (2007).
 - [36] “See Supplemental Material at [...], which includes details on derivations and Monte Carlo simulations as well as Refs. [45, 46].”
 - [37] T. Monz, P. Schindler, J. T. Barreiro, M. Chwalla,

- D. Nigg, W. A. Coish, M. Harlander, W. Hänsel, M. Henrich, and R. Blatt, Phys. Rev. Lett. **106**, 130506 (2011).
- [38] J. F. Goodwin, B. J. Brown, G. Stutter, H. Dale, R. C. Thompson, and T. Rudolph, arXiv:1407.1858 (2014).
- [39] S. T. Merkel, J. M. Gambetta, J. A. Smolin, S. Poletto, A. D. Córcoles, B. R. Johnson, C. A. Ryan, and M. Steffen, Phys. Rev. A **87**, 062119 (2013).
- [40] J. Medford, J. Beil, J. M. Taylor, S. D. Bartlett, A. C. Doherty, E. I. Rashba, D. P. DiVincenzo, H. Lu, A. C. Gossard, and C. M. Marcus, Nat. Nanotechnol. **8**, 654 (2013).
- [41] M. D. Shulman, S. P. Harvey, J. M. Nichol, S. D. Bartlett, A. C. Doherty, V. Umansky, and A. Yacoby, arXiv:1405.0485 (2014).
- [42] H. M. Wiseman and R. B. Killip, Phys. Rev. A **56**, 944 (1997).
- [43] K. Banaszek, G. M. D’Ariano, M. G. A. Paris, and M. F. Sacchi, Phys. Rev. A **61**, 010304 (1999).
- [44] A. G. Fowler, A. C. Whiteside, A. L. McInnes, and A. Rabbani, Phys. Rev. X **2**, 041003 (2012).
- [45] R. Y. Rubinstein and D. P. Kroese, *Simulation and the Monte Carlo method*, 2nd ed. (John Wiley & Sons, Hoboken, U.S.A., 2008) Chap. 2, pp. 51–54.
- [46] W. Press, B. Flannery, S. Teukolsky, and W. Vetterling, *Numerical Recipes in Fortran 77: the art of scientific computing*, 2nd ed. (Cambridge University Press, Cambridge, United Kingdom, 1992) Chap. 6, pp. 219–222.

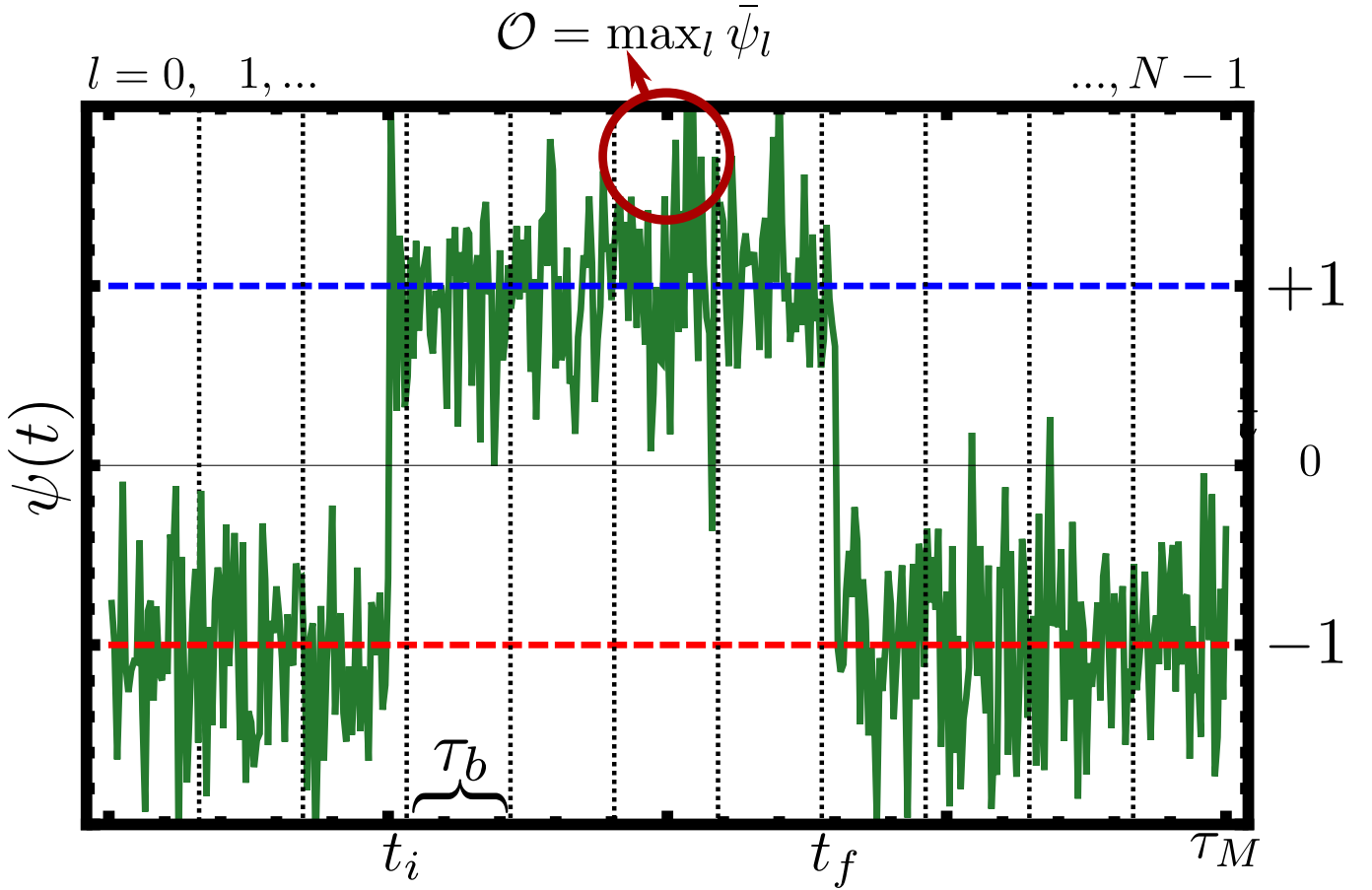


FIG. S.1. (Color online) Generic time-dependent signal $\psi(t)$ signalling the excited state $|+\rangle$ for the class of readouts discussed in Ref. [25]. The turn-on time t_i and pulse width $t_f - t_i$ both follow Poisson statistics. The measurement time τ_M is binned into subintervals of length τ_b and the observable \mathcal{O} is chosen to be the maximum of the signal over all bins. The parameters τ_M and τ_b are chosen to minimize the single-shot error rate.

SUPPLEMENTAL MATERIAL

REALISTIC NON-GAUSSIAN READOUT: THE PEAK-SIGNAL READOUT

To show that soft decoding of a readout leads to an advantage for realistic readouts, we have investigated the performance of soft decoding for the experimentally relevant [11, 24, 26] ‘peak-signal’ readout analyzed in Ref. [25]. In this section, we briefly summarize this readout for completeness.

In this readout, the ground and excited states $|-\rangle$ and $|+\rangle$ are mapped to a time-dependent signal $\psi(t)$ subject to Gaussian white noise. When the state is $|-\rangle$, the signal has a constant value -1 on average. When the state is $|+\rangle$, however, the signal is a square pulse starting at a random turn-on time t_i and ending at a random turn-off time t_f , as illustrated in Fig. S.1. The times t_i and $t_f - t_i$ each follow their own Poisson statistics. The measurement time τ_M is divided into N bins of length τ_b , with the average signal on the l^{th} bin being $\bar{\psi}_l$. The observable \mathcal{O} is then chosen to be the maximum of $\bar{\psi}_l$ over all bins. Finally, the measurement time τ_M , bin time τ_b and threshold ν are chosen to optimize the single-shot readout fidelity. Other choices for the observable \mathcal{O} have also been discussed in Refs. [25, 28].

Typical probability distributions $P(\mathcal{O}|\pm)$ for the peak-signal readout are shown in Fig. S.2 for two different values of the power signal-to-noise ratio r (integrated over a time $\langle t_f - t_i \rangle$). We see that the distributions have prominent non-Gaussian features. To perform fast sampling of these distributions, we first cut off the tails of each distribution (the lost probability weight is smaller than about 10^{-7}) and renormalize them. Next, we numerically integrate the analytic expressions for $P(\mathcal{O}|\pm)$ given in Ref. [25] to construct a linear interpolation of the inverse cumulative distribution function Q_{\pm}^{-1} associated with $P(\mathcal{O}|\pm)$, as shown in Fig. S.2. An independent sample of $P(\mathcal{O}|\pm)$ is then given by $\mathcal{O} = Q_{\pm}^{-1}(x)$, where x is a random number generated from a uniform distribution between 0 and 1 [45].

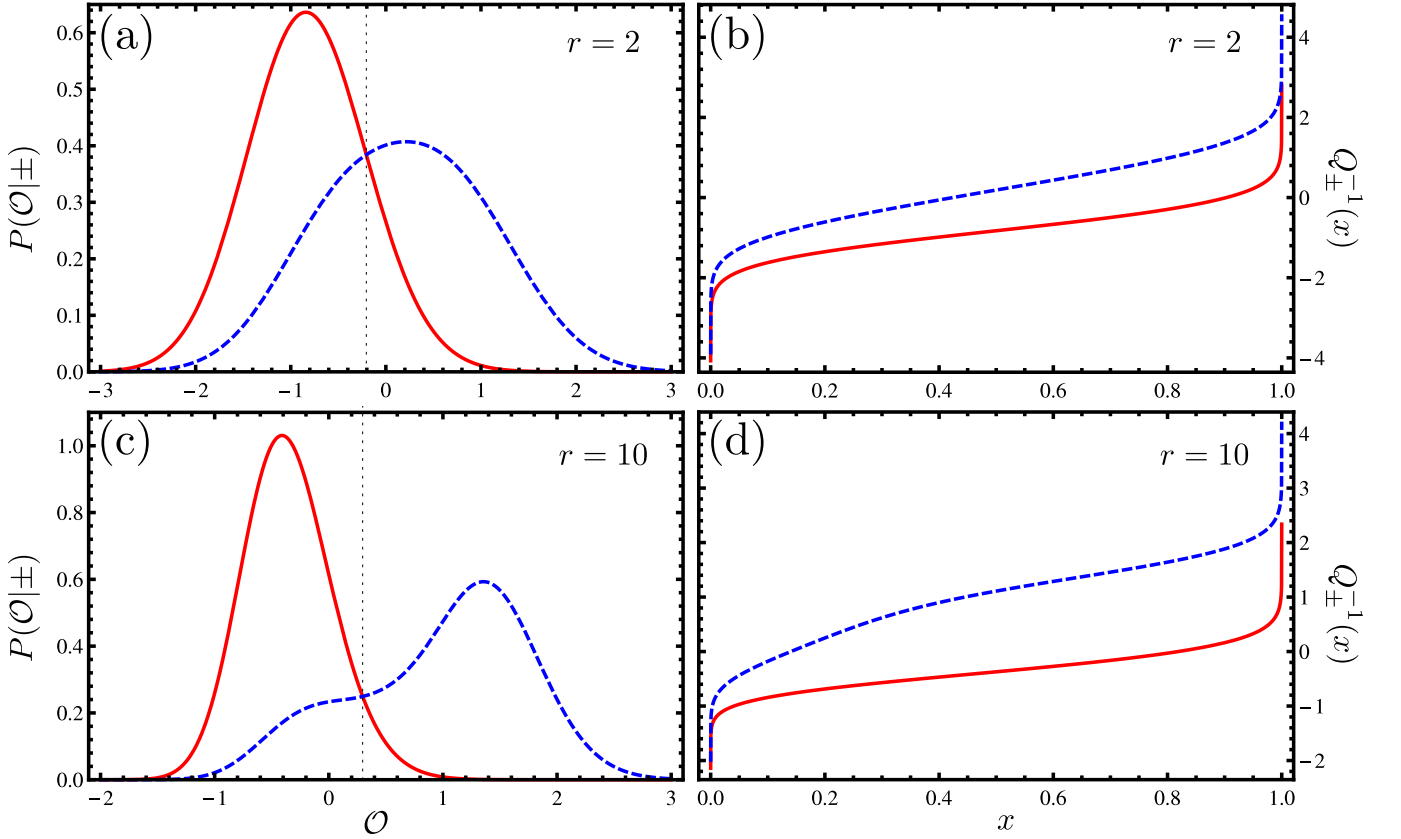


FIG. S.2. (Color online) Example of optimal peak-signal distributions $P(\mathcal{O}|\pm)$ given in Ref. [25] for a signal-to-noise ratio of (a) $r = 2$ and (c) $r = 10$. The dotted vertical line is the optimal threshold. In both cases, $\langle t_f - t_i \rangle / \langle t_i \rangle = 4$. The corresponding inverse cumulative distribution functions $Q_{\pm}^{-1}(x)$ used for fast sampling of the distributions are shown in panels (b) and (d).

ENHANCED STATE DETECTION WITH THE QUANTUM REPETITION CODE

Advantage of soft decoding

In this section, we give a brief derivation of the advantage obtained for enhanced state detection through soft decoding of the quantum repetition code for a Gaussian readout, Eq. (4) of the main text. Note that to benefit from enhanced state detection, coherent encoding of the logical state $|\psi\rangle = \alpha_0 |0\rangle + \alpha_1 |1\rangle$ into the state $|\psi_N\rangle = \alpha_0 |-\rangle^{\otimes N} + \alpha_1 |+\rangle^{\otimes N}$ is not necessary. For example, encoding into the mixed state $\rho_N = |\alpha_0|^2 |-\rangle\langle -|^{\otimes N} + |\alpha_1|^2 |+\rangle\langle +|^{\otimes N}$ (e.g. by allowing the qubits to purely dephase) gives the same advantage.

For the Gaussian readout, each qubit measurement yields an analog value \mathcal{O} conditioned on the qubit state $|\pm\rangle$ according to the probability distributions:

$$P(\mathcal{O}|\pm) = \sqrt{\frac{r}{2\pi}} e^{-\frac{(\mathcal{O} \mp 1)^2 r}{2}}. \quad (\text{S.1})$$

Here, r is the power signal-to-noise ratio and the signal is normalized to have average values ± 1 . If the qubit is read out in a single shot, each analog outcome \mathcal{O}_i is converted to a binary outcome $c_i = c_{\pm}$ by setting a threshold ν . The single-shot error rates conditional on the qubit state are:

$$\epsilon_- = P(c_+|-) = \int_{\nu}^{\infty} d\mathcal{O} P(\mathcal{O}|-), \quad \epsilon_+ = P(c_-|+) = \int_{-\infty}^{\nu} d\mathcal{O} P(\mathcal{O}|+). \quad (\text{S.2})$$

Assuming equal *a priori* probabilities for the ground and excited states, the average single-shot readout error rate

$\epsilon = (\epsilon_+ + \epsilon_-)/2$ is minimized by choosing $P(\nu|+) = P(\nu|-) \Rightarrow \nu = 0$. An explicit calculation of the integrals gives:

$$\epsilon = \epsilon_{\pm} = \frac{1}{2} \operatorname{erfc} \left(\sqrt{\frac{r}{2}} \right). \quad (\text{S.3})$$

Eq. (S.3) defines the binary symmetric readout associated with the Gaussian readout.

We assume for simplicity that when the N qubits of the quantum repetition code are measured, the N -qubits state collapses to $|+\rangle^{\otimes N}$ or $|-\rangle^{\otimes N}$ with equal probability. The resulting dataset consists of N analog readout outcomes \mathcal{O}_i . In the main text, we gave two likelihood ratios Λ_c and $\Lambda_{\mathcal{O}}$ for thresholded and analog readout outcomes, respectively. In both cases, if $\Lambda > 1$ ($\Lambda < 1$) we infer that the qubit state is $|1\rangle$ ($|0\rangle$). For the Gaussian readout, the likelihood ratios reduce to:

$$\Lambda_c = \left(\frac{1 - \epsilon}{\epsilon} \right)^{2n_+ - N}, \quad \Lambda_{\mathcal{O}} = \exp(2Nr\bar{\mathcal{O}}), \quad (\text{S.4})$$

where n_+ is the number of times that the outcome \mathcal{O}_i is converted to c_+ if the qubits are read out in a single shot and where $\bar{\mathcal{O}} = N^{-1} \sum_{i=1}^N \mathcal{O}_i$ is the sample average of the analog outcomes [7].

Since $\epsilon < 1/2$, the likelihood ratio Λ_c is equivalent to majority vote decoding of the repetition code. The corresponding average error rate ε_c is given by the probability that $n_+ > N/2$ given that $|0\rangle$ is encoded, which is the same as the probability that $n_+ < N/2$ given that $|1\rangle$ is encoded. If $N = 2M - 1$ is odd, ε_c is given by:

$$\varepsilon_c = \sum_{n_+ = \frac{N+1}{2}}^N \binom{N}{n_+} \epsilon^{n_+} (1 - \epsilon)^{N - n_+} = I_{\epsilon} \left(\frac{N+1}{2}, \frac{N+1}{2} \right). \quad (\text{S.5})$$

Here, $I_{\epsilon}(a, b)$ is the regularized incomplete beta function [46]. The error rate for $N = 2M$ is the same since the case $n_+ = N/2$ provides no information on the qubit state for a binary symmetric readout. For r large enough ($N\epsilon \ll 1$), Eq. (S.5) takes the approximate form:

$$\varepsilon_c \simeq \left(\frac{N}{\frac{N+1}{2}} \right) \frac{1}{(2\pi r)^{\frac{N+1}{4}}} e^{-\frac{(N+1)r}{4}}. \quad (\text{S.6})$$

Eq. (S.6) must be contrasted to the error rate for the likelihood ratio $\Lambda_{\mathcal{O}}$ in Eq. (S.4). The corresponding average error rate $\varepsilon_{\mathcal{O}}$ is given by the probability that $\bar{\mathcal{O}} > 0$ given that $|0\rangle$ is encoded, which is the same as the probability that $\bar{\mathcal{O}} < 0$ given that $|1\rangle$ is encoded. Since $P(\bar{\mathcal{O}}|1)$ and $P(\bar{\mathcal{O}}|0)$ are also Gaussians centered at ± 1 with signal-to-noise ratio Nr , the average error rate for the soft decoding of the readout apparatus is simply:

$$\varepsilon_{\mathcal{O}} = \int_0^{\infty} d\bar{\mathcal{O}} P(\bar{\mathcal{O}}|0) = \frac{1}{2} \operatorname{erfc} \left(\sqrt{\frac{Nr}{2}} \right). \quad (\text{S.7})$$

When $r \gg 1$, Eq. (S.7) becomes:

$$\varepsilon_{\mathcal{O}} \simeq \frac{1}{\sqrt{2\pi Nr}} e^{-\frac{Nr}{2}}. \quad (\text{S.8})$$

Inspection of Eqs. (S.6) and (S.8) suggests that $\varepsilon_{\mathcal{O}}$ decreases at approximately twice the rate of ε_c when N increases. Indeed, let N_c and $N_{\mathcal{O}}$ be the number of qubits required to achieve a target error rate ε , i.e. $\varepsilon_c(r, N_c) = \varepsilon_{\mathcal{O}}(r, N_{\mathcal{O}})$. Using Eqs. (S.5) and (S.7), we solve this equation for $N_{\mathcal{O}}$ to subleading order in $r \gg N_c$ and obtain the result discussed in the main text:

$$N_{\mathcal{O}} = \frac{N_c + 1}{2} + \frac{N_c - 1}{2} \frac{\ln r}{r} + O \left(\frac{N_c}{r} \right). \quad (\text{S.9})$$

This expression is valid for any odd $N_c \geq 1$. The asymptotic advantage, $N_{\mathcal{O}} \sim N_c/2$, has been discussed for arbitrary block codes in Ref. [?].

Encoding errors

In this section, we expand on the effect of encoding errors on the repetition code. As discussed in the main text, we consider only uncorrelated bit flip errors for simplicity. We consider both the Gaussian readout and the realistic peak-signal readout analyzed in Ref. [25] and summarized above.

Let η be the probability for any qubit of the code to flip during the encoding sequence. The likelihood ratio for analog readout outcomes takes the modified form:

$$\Lambda_{\mathcal{O}} \equiv \prod_{i=1}^N \Lambda_{\mathcal{O},i}, \quad (\text{S.10})$$

where $\Lambda_{\mathcal{O},i}$ is the likelihood ratio for a single qubit measurement:

$$\Lambda_{\mathcal{O},i} = \frac{P(\mathcal{O}_i|1)}{P(\mathcal{O}_i|0)} = \frac{(1-\eta)P(\mathcal{O}_i|+) + \eta P(\mathcal{O}_i|-)}{(1-\eta)P(\mathcal{O}_i|-) + \eta P(\mathcal{O}_i|+)}. \quad (\text{S.11})$$

If the *a priori* probabilities of the logical states $|0\rangle$ and $|1\rangle$ are equal, the single-shot threshold ν is obtained as usual from $P(\nu|1) = P(\nu|0) \Rightarrow P(\nu|+) = P(\nu|-)$. Therefore, sufficiently localized readout probability distributions such as the Gaussian distributions satisfy:

$$\Lambda_{\mathcal{O},i} \approx \frac{1-\eta}{\eta} \quad (\mathcal{O} \gg \nu), \quad \Lambda_{\mathcal{O},i} \approx \frac{\eta}{1-\eta} \quad (\mathcal{O} \ll \nu). \quad (\text{S.12})$$

Because $\Lambda_{\mathcal{O},i}$ is approximately constant above and below threshold, it seems that soft decoding of the analog readout outcomes is reduced to a simple thresholding procedure when η is finite. As illustrated in Fig. S.3, additional information can nevertheless be extracted from values \mathcal{O}_i falling near the threshold, where $\Lambda_{\mathcal{O},i}$ is non-constant, provided that η is small enough.

In the case of the Gaussian readout distributions, $\Lambda_{\mathcal{O},i}$ is non-constant for values of \mathcal{O} such that $\exp(-2r|\mathcal{O}|) \gtrsim \eta/(1-\eta)$, as illustrated in Fig. S.3. In order for a significant fraction of measured values to lie in that interval, we must have $|\mathcal{O}| \gtrsim 1$. Therefore:

$$\frac{\eta}{1-\eta} \lesssim e^{-2r} \Rightarrow \eta \lesssim \frac{e^{-2r}}{1+e^{-2r}}. \quad (\text{S.13})$$

In the limit $r \gg 1$, this reduces to $\eta \lesssim e^{-2r}$. Since Eq. (S.3) implies that $\epsilon \sim e^{-\frac{r}{2}}$ up to logarithmic corrections for $r \gg 1$, we conclude that η must be smaller than some power of ϵ . As shown in Fig. S.5, a similar upper bound on η exists for the non-Gaussian peak-signal readout of Ref. [25].

If each analog outcome \mathcal{O}_i is instead thresholded to a binary outcome $c_i = c_{\pm}$, the likelihood ratio is:

$$\Lambda_c = \left[\frac{(1-\eta)(1-\epsilon_+) + \eta\epsilon_-}{(1-\eta)\epsilon_- + \eta(1-\epsilon_+)} \right]^{n_+} \cdot \left[\frac{(1-\eta)\epsilon_+ + \eta(1-\epsilon_-)}{(1-\eta)(1-\epsilon_-) + \eta\epsilon_+} \right]^{N-n_+}, \quad (\text{S.14})$$

where n_+ is the number of qubits that are assigned the value c_+ . To show quantitatively that a significant advantage can be obtained by utilizing the analog readout outcomes in the presence of encoding errors, we must compare the performance of Eq. (S.14) to that of Eq. (S.10).

For both the Gaussian readout and the realistic peak-signal readout, we performed Monte-Carlo simulations of the error rates for maximum-likelihood decoding of the analog and thresholded readout outcomes. In both cases, we take the signal-to-noise ratio to be $r = 2$ and choose parameters that optimize the single-shot readout fidelity. We randomly choose the logical state $|0\rangle$ or $|1\rangle$ with equal probability and generate a random measurement record by sampling N independent values from the distributions $P(\mathcal{O}_i|1) = (1-\eta)P(\mathcal{O}_i|+) + \eta P(\mathcal{O}_i|-)$ or $P(\mathcal{O}_i|0) = (1-\eta)P(\mathcal{O}_i|-) + \eta P(\mathcal{O}_i|+)$, respectively. We then infer the state with both Eq. (S.10) and Eq. (S.14) and record an error if the decision is incorrect. We repeat the procedure 10^7 times (10^6 times) for the Gaussian readout (peak-signal readout) and obtain the error rate from the ratio of errors to the number of trials. The resulting error rates are shown in Figs. S.4 and S.5 respectively. For the Gaussian readout with $\eta = 0$, we instead plot the analytic expressions, Eqs. (S.5) and (S.7). Numerical values of the error rates for both readouts without encoding errors are tabulated in Table S.1 for convenience.

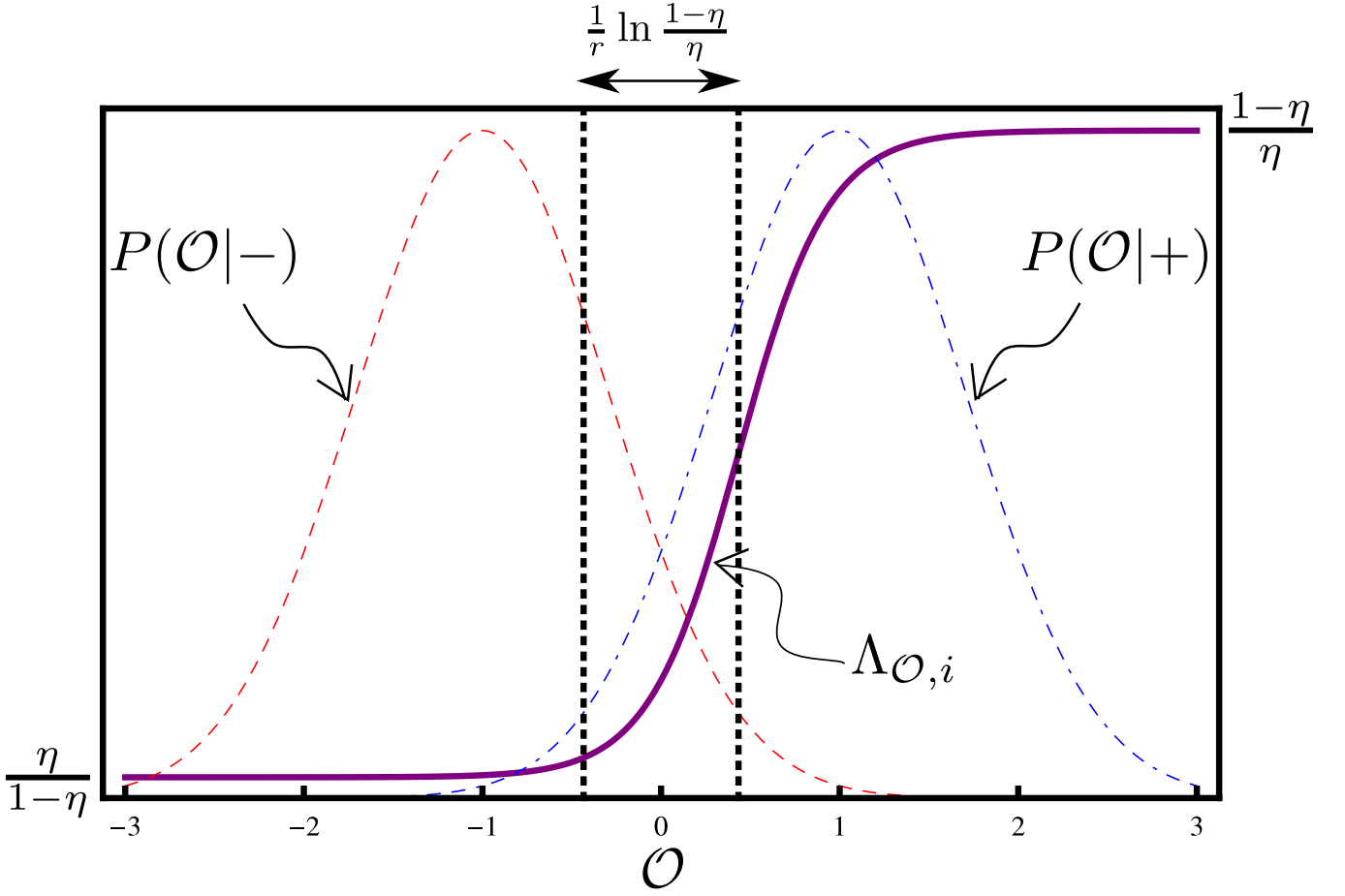


FIG. S.3. (Color online) Schematic plot of the likelihood ratio $\Lambda_{\mathcal{O},i}$ for a single qubit measurement, Eq. (S.11), as a function of \mathcal{O} for the Gaussian readout with $r = 2$ (the values on the vertical axis have been rescaled for clarity). Far above (below) threshold, $\Lambda_{\mathcal{O},i}$ is approximately constant. However, $\Lambda_{\mathcal{O},i}$ is non-constant on an interval of width $\frac{1}{r} \ln \frac{1-\eta}{\eta}$.

| Gaussian readout | | | | | | | | | |
|---------------------------------------|-------|-------|-------|--------|--------|--------|---------|---------|---------|
| N | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| $\varepsilon_c (10^{-2})$ | 7.86 | 7.86 | 1.76 | 1.76 | 0.431 | 0.431 | 0.110 | 0.110 | 0.0289 |
| $\varepsilon_{\mathcal{O}} (10^{-2})$ | 7.86 | 2.28 | 0.715 | 0.234 | 0.0783 | 0.0266 | 0.00914 | 0.00317 | 0.00110 |
| Peak-signal readout | | | | | | | | | |
| N | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| ε_c | 0.253 | 0.208 | 0.172 | 0.124 | 0.121 | 0.0845 | 0.0715 | 0.0630 | 0.0465 |
| $\varepsilon_{\mathcal{O}}$ | 0.253 | 0.166 | 0.116 | 0.0832 | 0.0608 | 0.0450 | 0.0334 | 0.0247 | 0.0187 |

TABLE S.1. Tabulated values of the thresholded and soft-decoded error rates ε_c and $\varepsilon_{\mathcal{O}}$ for different numbers N of repetition code qubits for both the Gaussian and peak-signal readouts. In both cases, the signal-to-noise ratio is $r = 2$ and there are no encoding errors, $\eta = 0$. These values correspond to those plotted in Figs. S.4 and S.5.

STATE AND PARAMETER ESTIMATION

In this section, we give a brief derivation of the asymptotic mean squared error of the maximum-likelihood estimator for $s_0 = \langle \sigma_z \rangle$ when applied to analog and thresholded readout outcomes. We also review the soft average discussed in Ref. [29]. In all cases, we estimate s_0 with N independent analog or thresholded readout outcomes, $\{\mathcal{O}_i\}$ or $\{c_i\}$,

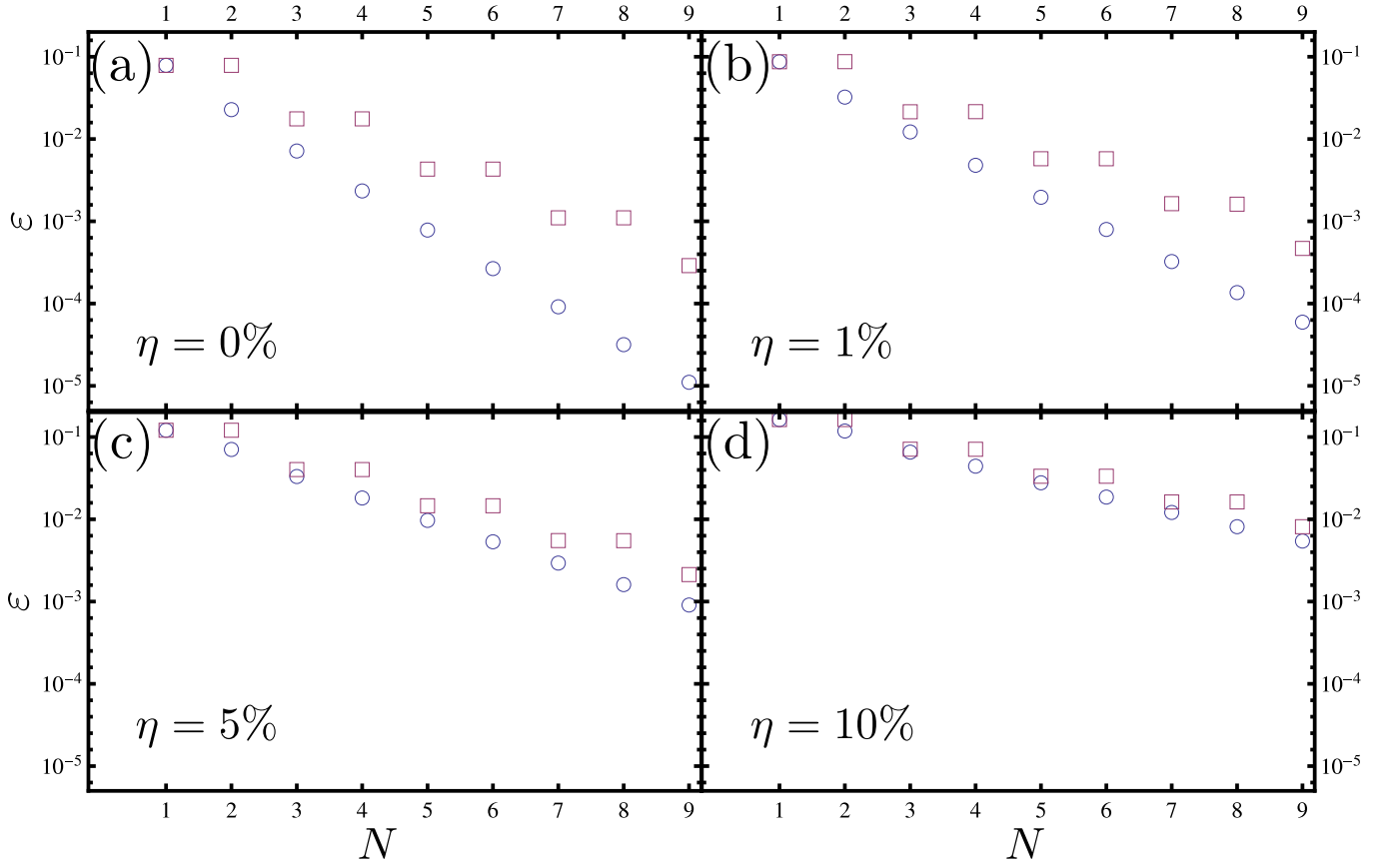


FIG. S.4. (Color online) Simulated repetition code error rates ε as a function of the number of qubits N for the Gaussian readout with signal-to-noise ratio $r = 2$. The error rate was obtained for soft decoding (blue circle), Eq. (S.10), and thresholding (magenta square), Eq. (S.14), of the readout outcomes. The error rate for thresholding is the same for $N = 2M$ as for $N = 2M - 1$ since the case $n_+ = N/2$ provides no information on the qubit state ($\epsilon_+ = \epsilon_- = \epsilon$). Each panel corresponds to a different encoding error rate η . For $\eta = 0$, we plotted Eqs. (S.5) and (S.7). For $\eta \neq 0$, the error rates were calculated by generating 10^7 random measurement records $\{\mathcal{O}_i\}$ sampled with equal probability from $P(\{\mathcal{O}_i\} | 1)$ and $P(\{\mathcal{O}_i\} | 0)$.

following a distribution of the form:

$$P(\mathcal{O}_i/c_i|s_0) = \frac{1+s_0}{2}P(\mathcal{O}_i/c_i|+) + \frac{1-s_0}{2}P(\mathcal{O}_i/c_i|-). \quad (\text{S.15})$$

In the following, we will denote statistical expectation values with respect to Eq. (S.15) by the double brackets $\langle\langle \cdot \rangle\rangle$. The maximum-likelihood estimator is the value s that maximizes the log-likelihood function:

$$\ell(s) = \frac{1}{N} \sum_{i=1}^N \ln P(\mathcal{O}_i/c_i|s), \quad (\text{S.16})$$

under the constraint $-1 \leq s \leq 1$.

Thresholded readout outcomes

First we assume that the values \mathcal{O}_i are thresholded to a binary outcome c_{\pm} , where the threshold ν is chosen to satisfy $P(\nu|+) = P(\nu|-)$.

To obtain the maximum-likelihood estimator, we must maximize the likelihood function, Eq. (S.16). We first note that Bayes' rule gives the probability of an outcome c_i given the true expectation s_0 :

$$P(c_i|s_0) = \frac{1+s_0}{2}P(c_i|+) + \frac{1-s_0}{2}P(c_i|-). \quad (\text{S.17})$$

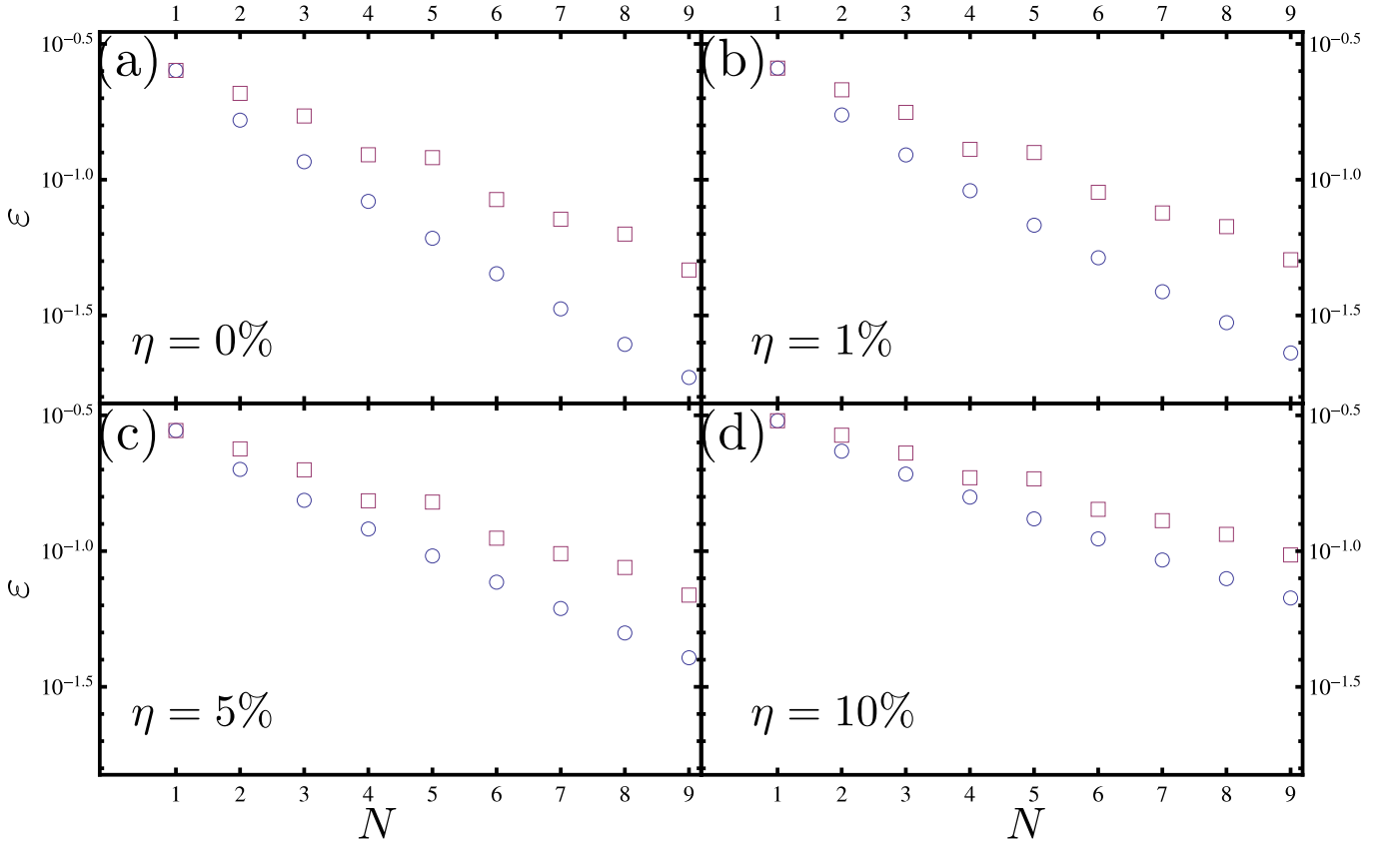


FIG. S.5. (Color online) Simulated repetition code error rates ε as a function of the number of qubits N for the peak-signal readout of Fig. S.1 with signal-to-noise ratio $r = 2$ and $\langle t_f - t_i \rangle / \langle t_i \rangle = 4$. The error rate was obtained for soft decoding (blue circle), Eq. (S.10), and thresholding (magenta square), Eq. (S.14), of the readout outcomes. The error rate for thresholding follows a jagged pattern since the case $n_+ = N/2$ only gives partial information on the qubit state ($\epsilon_+ \neq \epsilon_-$). Each panel corresponds to a different encoding error rate η . The error rates were calculated by generating 10^6 random measurement records $\{\mathcal{O}_i\}$ sampled with equal probability from $P(\{\mathcal{O}_i\} | 1)$ and $P(\{\mathcal{O}_i\} | 0)$.

Here, the transition probabilities of the binary readout are given by the conditional single-shot error rates:

$$P(c_- | +) \equiv \epsilon_+ = \int_{-\infty}^{\nu} d\mathcal{O} P(\mathcal{O} | +), \quad P(c_+ | -) \equiv \epsilon_- = \int_{\nu}^{\infty} d\mathcal{O} P(\mathcal{O} | -). \quad (\text{S.18})$$

Thus, Eq. (S.17) becomes:

$$P(c_+ | s_0) = \frac{1 + s_0}{2} (1 - \epsilon_+) + \frac{1 - s_0}{2} \epsilon_-, \quad P(c_- | s_0) = \frac{1 + s_0}{2} \epsilon_+ + \frac{1 - s_0}{2} (1 - \epsilon_-). \quad (\text{S.19})$$

Next, we use the form of Eq. (S.19) in the log-likelihood function, Eq. (S.16), and optimize with respect to s . Maximizing without the constraint $-1 \leq s \leq 1$ (i.e. setting $d\ell(s)/ds = 0$), the optimum is the thresholded average:

$$s_{\text{TA}} = \frac{1}{N} \sum_{i=1}^N c_i, \quad (\text{S.20})$$

where the binary outcomes $c_i = c_{\pm}$ are chosen to be:

$$c_+ = \frac{1 + (\epsilon_+ - \epsilon_-)}{1 - (\epsilon_+ + \epsilon_-)}, \quad c_- = -\frac{1 - (\epsilon_+ - \epsilon_-)}{1 - (\epsilon_+ + \epsilon_-)}. \quad (\text{S.21})$$

In the limit of large N , the estimate is unlikely to fall outside the region $-1 \leq s \leq 1$. In this asymptotic limit, the

estimate is unbiased:

$$\langle\langle s_{\text{TA}} \rangle\rangle = P(c_+|s_0)c_+ + P(c_-|s_0)c_- \quad (\text{S.22})$$

$$= \frac{1+s_0}{2} [(1-\epsilon_+)c_+ + \epsilon_+c_-] + \frac{1-s_0}{2} [\epsilon_-c_+ + (1-\epsilon_-)c_-] = s_0. \quad (\text{S.23})$$

In this case, the asymptotic mean squared error ζ_{TA} of the maximum-likelihood estimate, Eq. (S.20), is equal to its asymptotic variance and is given by the central limit theorem:

$$\zeta_{\text{TA}} = \langle\langle \Delta s_{\text{TA}}^2 \rangle\rangle \sim \frac{\langle\langle \Delta c^2 \rangle\rangle}{N} = \frac{P(c_+|s_0)c_+^2 + P(c_-|s_0)c_-^2 - s_0^2}{N}. \quad (\text{S.24})$$

In the special case of a binary symmetric readout with $\epsilon_+ = \epsilon_- = \epsilon$, we have $c_+ = -c_- = (1-2\epsilon)^{-1}$ and we recover the expression given in Ref. [29]:

$$\zeta_{\text{TA}} \sim \frac{(1-2\epsilon)^{-2} - s_0^2}{N}. \quad (\text{S.25})$$

Analog readout outcomes

The asymptotic mean squared error ζ_{SD} of the maximum-likelihood estimator applied to the analog readout outcomes is equal to its asymptotic variance, which saturates the Cramér-Rao bound [30]:

$$\zeta_{\text{SD}} \sim \frac{1}{NF(s_0)}, \quad (\text{S.26})$$

where $F(s_0)$ is the Fisher information of the distribution (S.15):

$$F(s_0) = \left\langle\left\langle \left(\frac{\partial \ln P(\mathcal{O}|s_0)}{\partial s_0} \right)^2 \right\rangle\right\rangle = - \left\langle\left\langle \frac{\partial^2 \ln P(\mathcal{O}|s_0)}{\partial s_0^2} \right\rangle\right\rangle. \quad (\text{S.27})$$

The last equality in Eq. (S.27) is obtained through integration by parts. Differentiating Eq. (S.15) twice gives an explicit form for $F(s_0)$:

$$F(s_0) = \frac{1}{4} \int d\mathcal{O} \frac{[P(\mathcal{O}|+) - P(\mathcal{O}|-)]^2}{P(\mathcal{O}|s_0)}. \quad (\text{S.28})$$

Expanding the integrand, we have:

$$F(s_0) = \frac{1}{4} \left[\int d\mathcal{O} \frac{P(\mathcal{O}|+)^2}{P(\mathcal{O}|s_0)} + \int d\mathcal{O} \frac{P(\mathcal{O}|-)^2}{P(\mathcal{O}|s_0)} - 2 \int d\mathcal{O} \frac{P(\mathcal{O}|+)P(\mathcal{O}|-)}{P(\mathcal{O}|s_0)} \right]. \quad (\text{S.29})$$

When the readout distributions $P(\mathcal{O}|\pm)$ are very well-separated, the Fisher information only contains the shot noise contribution $F(s_0) = 1/(1-s_0^2)$. We isolate this contribution in Eq. (S.29) and upon simplification we find:

$$F(s_0) = \frac{1}{1-s_0^2} - \frac{1}{1-s_0^2} I, \quad I = \int d\mathcal{O} \frac{P(\mathcal{O}|+)P(\mathcal{O}|-)}{P(\mathcal{O}|s_0)}, \quad (\text{S.30})$$

where I is an overlap integral containing all information about the intrinsic measurement noise described by $P(\mathcal{O}|\pm)$. Therefore, the asymptotic mean squared error of the maximum-likelihood estimator applied to the analog readout outcomes is:

$$\zeta_{\text{SD}} \sim \frac{1-s_0^2}{1-I}. \quad (\text{S.31})$$

Bias-corrected soft average

Another possible estimator for the qubit expectation value is the soft average discussed in Ref. [29]:

$$s_{\text{SA}} = \frac{1}{N} \sum_{i=1}^N \mathcal{O}_i. \quad (\text{S.32})$$

We compare the performance of this estimator to the previously discussed estimators, s_{TA} and s_{SD} , for completeness. The expectation value of Eq. (S.32) with respect to $P(\mathcal{O}|s_0)$ has the form:

$$\langle\langle s_{\text{SA}} \rangle\rangle = As_0 + B, \quad (\text{S.33})$$

where:

$$A = \frac{\langle\langle \mathcal{O} \rangle\rangle_+ - \langle\langle \mathcal{O} \rangle\rangle_-}{2}, \quad B = \frac{\langle\langle \mathcal{O} \rangle\rangle_+ + \langle\langle \mathcal{O} \rangle\rangle_-}{2}. \quad (\text{S.34})$$

Here, we define the conditional expectations $\langle\langle \mathcal{O} \rangle\rangle_{\pm} = \int d\mathcal{O} P(\mathcal{O}|\pm)\mathcal{O}$. Thus, the soft average of Eq. (S.32) is biased for general readout probability distributions $P(\mathcal{O}|\pm)$.

To obtain an unbiased estimate, we replace Eq. (S.32) by the soft average of the rescaled values $\mathcal{O}'_i = (\mathcal{O}_i - B)/A$:

$$s_{\text{SA}} = \frac{1}{N} \sum_{i=1}^N \mathcal{O}'_i = \frac{1}{N} \sum_{i=1}^N \frac{\mathcal{O}_i - B}{A}. \quad (\text{S.35})$$

The asymptotic mean squared error ζ_{SA} of the unbiased soft average estimate, Eq. (S.35), is equal to its asymptotic variance and is given by the central limit theorem:

$$\zeta_{\text{SA}} = \langle\langle \Delta s_{\text{SA}}^2 \rangle\rangle = \frac{\Delta \mathcal{O}'^2}{N} = \frac{\langle\langle \mathcal{O}'^2 \rangle\rangle - s_0^2}{N}. \quad (\text{S.36})$$

In terms of the original observable \mathcal{O} , this becomes:

$$\zeta_{\text{SA}} = \frac{\Delta \mathcal{O}^2}{A^2 N} = \frac{\langle\langle \mathcal{O}^2 \rangle\rangle - (As_0 + B)^2}{A^2 N}. \quad (\text{S.37})$$

In the special case of the Gaussian readout, Eq. (S.1), we have $A = 1$ and $B = 0$. Direct calculation of $\langle\langle \mathcal{O}^2 \rangle\rangle$ then yields the result given in Ref. [29]:

$$\zeta_{\text{SA}} = \frac{1 + r^{-1} - s_0^2}{N}. \quad (\text{S.38})$$

Fig. 6 compares the asymptotic performance of the soft average s_{SA} to that of the maximum-likelihood estimates s_{TA} and s_{SD} as a function of the signal-to-noise ratio r , for both the Gaussian and the peak-signal readouts. As noted in Ref. [29], the soft average outperforms the thresholded average s_{TA} for low r . This is because the distribution $P(\mathcal{O}|s_0)$ approaches a Gaussian centered at s_0 when $r \rightarrow 0$ for both readouts, $P(\mathcal{O}|s_0) \simeq \sqrt{\frac{r}{2\pi}} e^{-\frac{(\mathcal{O}-s_0)^2 r}{2}}$, and the maximum-likelihood estimator for the mean of a Gaussian coincides with the soft average. In that case, the soft average s_{SA} is therefore the same as the soft-decoded estimate s_{SD} . However, the soft average estimate offers suboptimal performance for finite r and suffers from a significant loss in performance compared to s_{TA} and s_{SD} when r becomes large. In contrast, the soft-decoded estimate s_{SD} is optimal for all r .

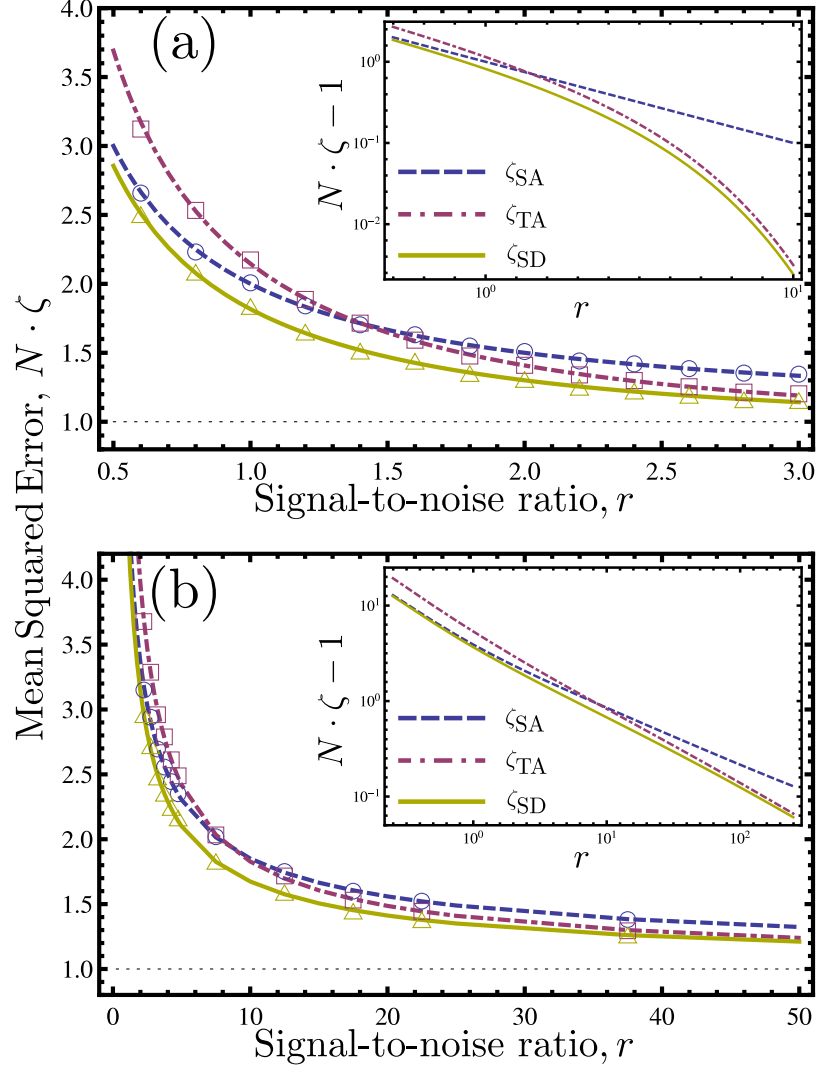


FIG. S.6. (Color online) Comparison of the asymptotic mean squared error of the soft average s_{SA} (dashed blue), Eq. (S.37), to that of the maximum-likelihood estimates s_{TA} (dot-dashed magenta), Eq. (S.24), and s_{SD} (solid gold), Eq. (S.31), for (a) the Gaussian readout and (b) the peak-signal readout. The finite- N MSEs for the soft average s_{SA} (blue circles), the thresholded average s_{TA} (magenta squares) and the soft-decoded estimate s_{SD} (gold triangles) are obtained from 5×10^4 randomly generated measurement records with $N = 100$. Inset: Asymptotic MSEs on a logarithmic scale.

-
- [1] J. M. Elzerman, R. Hanson, L. H. W. Van Beveren, B. Witkamp, L. M. K. Vandersypen, and L. P. Kouwenhoven, *Nature* **430**, 431 (2004).
 - [2] A. Morello, J. J. Pla, F. A. Zwanenburg, K. W. Chan, K. Y. Tan, H. Huebl, M. Möttönen, C. D. Nugroho, C. Yang, J. A. van Donkelaar, *et al.*, *Nature* **467**, 687 (2010).
 - [3] M. Veldhorst, J. C. C. Hwang, C. H. Yang, A. W. Leenstra, B. de Ronde, J. P. Dehollain, J. T. Muhonen, F. E. Hudson, K. M. Itoh, A. Morello, and A. S. Dzurak, arXiv preprint arXiv:1407.1950 (2014).
 - [4] B. D’Anjou and W. A. Coish, *Phys. Rev. A* **89**, 012313 (2014).
 - [5] J. Gambetta, W. A. Braff, A. Wallraff, S. M. Girvin, and R. J. Schoelkopf, *Phys. Rev. A* **76**, 012325 (2007).
 - [6] R. Y. Rubinstein and D. P. Kroese, *Simulation and the Monte Carlo method*, 2nd ed. (John Wiley & Sons, Hoboken, U.S.A., 2008) Chap. 2, pp. 51–54.
 - [7] In Ref. [12], a two-qubit repetition code was implemented using two trapped ions. The enhanced detection was obtained by collecting the total fluorescence $\bar{\mathcal{O}}$ of both ions. We note that according to Eq. (S.4), this effectively implements soft decoding of the repetition code if the fluorescence counts follow a Gaussian readout distribution. For a general readout, however, knowledge of $\bar{\mathcal{O}}$ is not sufficient to determine $\Lambda_{\mathcal{O}}$; in this case the analog observable \mathcal{O}_i must be recorded for each qubit.
 - [8] W. Press, B. Flannery, S. Teukolsky, and W. Vetterling, *Numerical Recipes in Fortran 77: the art of scientific computing*, 2nd ed. (Cambridge University Press, Cambridge, United Kingdom, 1992) Chap. 6, pp. 219–222.
 - [9] D. Chase, *Information Theory*, *IEEE Transactions on* **18**, 170 (1972).
 - [10] C. A. Ryan, B. R. Johnson, J. M. Gambetta, J. M. Chow, M. P. da Silva, O. E. Dial, and T. A. Ohki, arXiv:1310.6448 (2013).
 - [11] H. Cramér, *Mathematical methods of statistics* (Princeton University Press, Princeton, NJ, 1946) Chap. 32-33.
 - [12] T. Schaetz, M. D. Barrett, D. Leibfried, J. Britton, J. Chiaverini, W. M. Itano, J. D. Jost, E. Knill, C. Langer, and D. J. Wineland, *Phys. Rev. Lett.* **94**, 010501 (2005).